Journal of Surgical Research 123, 268–274 (2005)
doi:10.1016/j.jss.2004.08.014

Measuring Medical Education and Curriculum During Orthopedic Surgical Residency[1]

John F. Kragh, Jr., LTC(P), MC, USA,*,[2] Mark R. Bagg, COL, MC, USA,* John A. Ward, PhD,† George E. Omer, Jr., COL(RET.), MC, USA‡

*Orthopaedic Service Brooke Army Medical Center, and †Department of Clinical Investigation, Brooke Army Medical Center, Fort Sam Houston, Texas; and ‡Department of Orthopaedics and Rehabilitation, University of New Mexico, Albuquerque, New Mexico

Submitted for publication June 7, 2004

Background. Evaluating residency programs requires objective assessment tools, but few are readily available. The purpose of this study was to measure education by correlating resident test scores with several measurements of educator performance.

Materials and methods. The study group included residents and educators from a single residency program. We performed a retrospective analysis of scores from the Orthopaedic In-Training Examination collected during a 6-year period. Resident examination scores were indexed by dividing program averages by national averages to determine yearly score trends and then were correlated with educator attendance and teaching hours. Subspecialty scores were ranked to gauge residency strengths and weaknesses. Teaching hours devoted to subspecialties were compared with test scores to measure curricular emphases and to appraise teaching efficiency.

Results. Yearly average examination scores were proportional to national averages ($P < 0.001$). However, of 3436 possible educator-score associations, only 15 scores correlated highly ($r > 0.9$) with educators, and only 26 were significant ($P < 0.05$). Trend analysis put subspecialty scores in yearly perspective. Ranking was inaccurate until scores were indexed to the national average. In 2002, the distribution of 238 teaching hours ranged from 4 to 48 h for subspecialties, and 9 of 12 subspecialties were emphasized disproportionately to the examination. Teaching efficiency varied more than 10-fold by subspecialty.

Conclusions. The creation of a score index helped to identify and address imbalances between teaching hours devoted to subspecialties and resident needs as evidenced by low In-Training examination scores. The present study improved educator accountability by correlating measurements of teaching and learning.

Key Words: graduate medical education measurement; residency; surgery; assessment; curriculum; academic medicine.

INTRODUCTION

Residency education affects good clinical practice, and regular assessment helps check satisfactory residency education [1–4]. Educators seek useful assessment tools [4, 5], but few specific assessment methods are readily available [4, 6]. For example, tools offered by the Accreditation Council of Graduate Medical Education focus on individual resident assessment [7], but we found these tools problematic for residency program assessment. For example, the assessment of individuals helped educators appraise individuals but did not help educators manage program-as-a-whole issues, like the quantities of curriculum by subspecialty. The need to assess residency education with objective criteria was plain yet unfulfilled [3–5] so tools were devised from available data to quantify education. We derived assessment tools from the Orthopaedic In-Training Examination (OITE), an evaluation of overall learning within 12 subspecialties used for medical knowledge assessment since 1963 [8–11]. The written examination, in 12 clinical topics, averages 265 total

[1] The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or reflecting the views of the Department of Defense or United States Government. The authors are employees of the U.S. government. This work was prepared as part of their official duties and, as such, there is no copyright to be transferred.

[2] To whom correspondence and reprint requests should be addressed at Extremity Trauma Study Branch, US Army Institute of Surgical Research, 3600 Chambers Ave, Bldg 3611 Rm 292-1, Ft Sam Houston, TX 78234-6315. E-mail: john.kragh@amedd.army.mil.

0022-4804/05 $30.00

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302 Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number

1. REPORT DATE **01 FEB 2005**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**

4. TITLE AND SUBTITLE **Measuring medical education and curriculum during orthopedic surgical residency**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) **Kragh J. F. Jr., Bagg M. R., Ward J. A., Omer G. E. Jr.,**	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **United States Army Institute of Surgical Research, JBSA Fort Sam Houston, TX 78234**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a REPORT **unclassified**	b ABSTRACT **unclassified**	c THIS PAGE **unclassified**	**SAR**	**7**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

questions, and lasts up to 5 h as taken by each resident [12]. The OITE, sponsored by the American Academy of Orthopaedic Surgeons, is a valid indicator of medical knowledge [13]. Scorers from the American Academy of Orthopaedic Surgeons report test results as number of subspecialty questions, and the program and national averages of correct answers by subspecialty [14].

The aim of the present study was to measure teaching and learning to identify areas in need of redress. We measured the success of our residency by correlating educator attendance with resident OITE scores, measured learning by yearly scoring trends, ranked subspecialty scores, counted teaching hours, compared teaching hours and program scores, and calculated efficiency of teaching.

MATERIALS AND METHODS

Study Design

We designed a retrospective study to analyze data gathered from 1997 to 2002 to test the hypothesis that there was a difference in OITE scores as a function of educator attendance. The OITE is divided by subspecialties such as Spine, Hand, Hip and Knee Reconstruction, etc., or by general categories, such as Orthopaedic Diseases, including musculoskeletal tumors. We use the term subspecialty generally to refer to all test categories in the present study. The table contains definitions of study terms. The study group averaged 10 staff surgeons and 20 residents from a single residency.

Program Scores, National Scores, and the Score Index

We averaged resident scores from the OITE to get annual program average by subspecialty. Program averages were dependent variables. We plotted the program average (number of correct answers divided by subspecialty questions) and used linear regression to determine the relationship between program and national scores (program average = 0.85 × national average +0.11, $P < 0.001$, $r = 0.855$, $r^2 = 0.73$). To account for variability in yearly scores due to the difficulty of subspecialty topics, variations in the number of questions between tests, and to permit comparisons among subspecialties, we created a score index. The score index for each subspecialty was obtained by dividing the program average by the national average and constituted another dependent variable to evaluate program and educator effectiveness by subspecialty.

Educator Associations with Resident Examination Scores

Educator attendance, the presence or absence of individual surgeons within the program, was an independent variable. A surgeon had to be in the program for four or more months prior to the examination to qualify as in attendance. For example, if a surgeon was deployed to war for more than 8 months, then he was "absent" as an educator regarding that year's examination. Surgeon attendance (presence or absence) was transformed to a scale variable (1 or 0, respectively). Surgeon attendance was correlated with subspecialty program averages and score index. We used correlations to screen the data to determine which relationships needed further analysis by score trends over time.

Examination Score Trends

We charted subspecialty score trends over time to compare educator attendance with the score index and national and program scores. Differences of less than 10% between program and national averages were considered small; differences of 10% or more were considered large. If correlation and graphical analysis showed a relationship between surgeon attendance and scores within the surgeon's subspecialty, then we confirmed the relationship with analysis of variance.

TABLE 1

Definitions of Educational Terms

Term	Definition
Program average	For the residency, the number of correct OITE answers divided by the total number of questions per subspecialty, expressed as a percent.
National average	Nationwide, the number of correct OITE answers divided by the total number of OITE questions per subspecialty, expressed as a percent.
Score index	Program average divided by the national average.
Teaching hours	Residency teaching hours per subspecialty divided by total teaching hours for all subspecialties, expressed as a percent.
Emphasis	Teaching hours divided by percentage of OITE questions per subspecialty.
Efficiency	Score index expressed as a percent divided by percent teaching hours per subspecialty.

Subspecialty Score Rankings

We ranked subspecialties subjectively and objectively by the score index (program average divided by national average) to see how well we assessed high and low scores. We ranked subspecialties subjectively by surveying 12 residency personnel (7 residents and 5 educators) in 2002. The survey listed the 12 subspecialties and asked respondents to rank subspecialty performance for the program on the 2002 test. Ranked 1 to 12 from highest to lowest, subspecialty average ranks were compared to ranked score index (program average divided by national average) for 2002. Results were displayed graphically.

Teaching Hours, Emphasis, and Efficiency

We counted educational hours in 1-h blocks from the 2002 curriculum schedule to quantify subspecialty teaching effort devoted to events. Events included lectures, conferences, workshops, grand rounds, and symposia. We included only events that ten or more residents attended as a group. Resident availability averaged 11 residents during the study period because of external rotations. At least one attending participated in all educational events except for some resident group study sessions. Forty-eight hours of group study were not subspecialty-specific and were not counted. Teaching hours were considered as percentages (number of subspecialty hours divided by total teaching hours; Table 1).

We wanted to see how teaching hours were balanced with the examination, and so we calculated emphasis (percent teaching hours divided by the percentage of subspecialty questions). The benchmark emphasis was 1, and we display the results graphically. High emphasis was above 1.5, and low emphasis was below 0.5. To assess a trait of teaching hours we calculated efficiency (score index expressed as a percent divided by percent teaching hours) and displayed the results graphically.

To determine the relationship between teaching hours or effort devoted to subspecialties and the OITE emphasis on subspecialties, we plotted subspecialty teaching hours as a function of the percent-

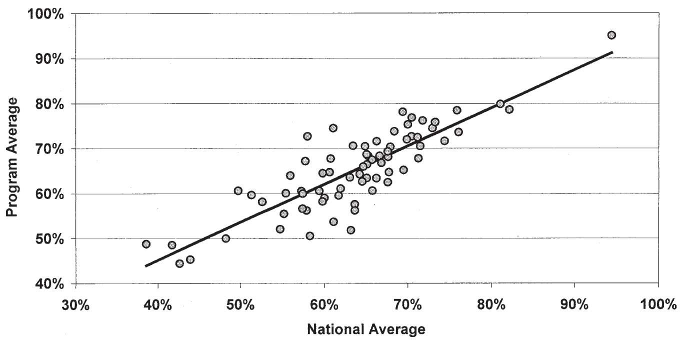

FIG. 1. OITE, program *versus* national averages. The linear regression line is displayed. Twelve subspecialties during the course of 6 years created a total of 72 data points. Program averages were directly proportional to national averages ($P < 0.001$).

age of number of subspecialty questions from 2002 and performed linear regression (teaching hours = 0.109 × questions +0.074, P = 0.783, r = 0.089). We analyzed similarly the subspecialty program scores and teaching hours (program score = 0.075 × teaching hours +0.058, P = 0.708, r = 0.326), and score index and teaching hours (score index = 0.404 × teaching hours +1.051, P = 0.299, r = 0.121) when score index and teaching hours were expressed as percentages.

Statistical Analysis

We used Pearson's product-moment correlation coefficient to compare resident scores to surgeon attendance. We counted high ($r > 0.9$) and significant ($P < 0.05$) correlations and then performed two-tailed t tests. Significant associations ($P < 0.05$) were positive or negative. A positive association was an increased subspecialty score with surgeon attendance while a negative association was a decreased score. We used Microsoft Office Professional 97 for data management (Microsoft Inc., Redmond, WA) and SPSS, version 11.5 (SPSS Inc., Chicago, IL) for statistical analysis.

RESULTS

Program scores were directly proportional to national scores (Fig. 1). The result was significant ($P < 0.001$), indicating that the difficulty of individual subspecialty testing itself was a major factor in determining scores. Points greater than 2 standard errors from the line were 10% different than the expected value; this finding indicated a score index (program average divided by national average) may be a better assessment tool. Score index results adjust for differences in difficulty of subspecialty testing and provided additional perspective to program scores. For an example from Fig. 1, the lowest (far left; 49% program average, 39% national average, 1.26 score index) and highest (far right; 95%, 94%, 1.01) data points were without a clear relationship for program and national averages, but the score index made the results more clear. The r^2 value of 0.73 indicated that 73% of the variability can be explained by national average; in other words the difficulty of the test was a dominant factor in determining mean resident scores in our program.

Educator Associations With Resident Scores

Calculations for 24 attending surgeons and 12 subspecialties during the course of 6 years for 2 score measures—program averages and score index results—created 3456 associations. The program average had three high ($r > 0.9$) and eight significant ($P < 0.05$) correlations with surgeons. The score index had 12 high and 18 significant correlations with surgeons. Most correlations appeared coincidental on graphical analysis, but one program average correlation for Sports Medicine ($r = -0.960$) with surgeon Z, was significant ($P = 0.002$) and the effect was large (Fig. 2A). Surgeon Z was not a sports medicine surgeon and was present only in 1999.

Score Trends

We displayed subspecialty score trends by year (Fig. 2A and B). In 2002 for Orthopaedic Sciences, the program average was 22% above the national average (Fig. 2B). Surgeons B and U were available in 2002 only, whereas surgeons I and M were available in all years except 2002. These associations were significant for the score index ($P = 0.002$) but not for program score ($P > 0.05$).

Subspecialty Score Rankings

The list of 2002 subjective subspecialty rankings differed from the objective ranking with five subspecialties differing by four to eight positions. Highest ranked score index results were in Orthopaedic Science, Orthopaedic Diseases, Hand, and Rehabilitation; lowest ranked score index results were in Sports Medicine, Pediatric Orthopaedics, Hip and Knee Reconstruction, and Spine in descending order (Fig. 3). The worst three score index results on both 2002 and 6-year rankings were the same (i.e., Pediatric Orthopaedics, Spine, and Hip and Knee Reconstruction), but only Pediatric Orthopaedics was subjectively ranked in the bottom three positions. Subjective ranks for five subspecialties were misleading in that they were different than objective ranks. Radial graphing of objective rankings indicated relative scores precisely for every subspecialty (Fig. 3). The general improvement from the 6-year average to the 2002 scores may be from increased educator interest in academics.

Analysis of Teaching Hours, Emphasis, and Efficiency

For teaching hours, the distribution of 238 educational hours ranged from 1.7% (4/238) for Orthopaedic Diseases to 20.2% (48/238) for Orthopaedic Science (Fig. 4A). The most teaching hours were devoted to Orthopaedic Science, Trauma, and Hand; the least hours to Rehabilitation, Sports Medicine, and Orthopaedic Diseases in descending order. Orthopaedic Science was the only topic systematically delegated to all

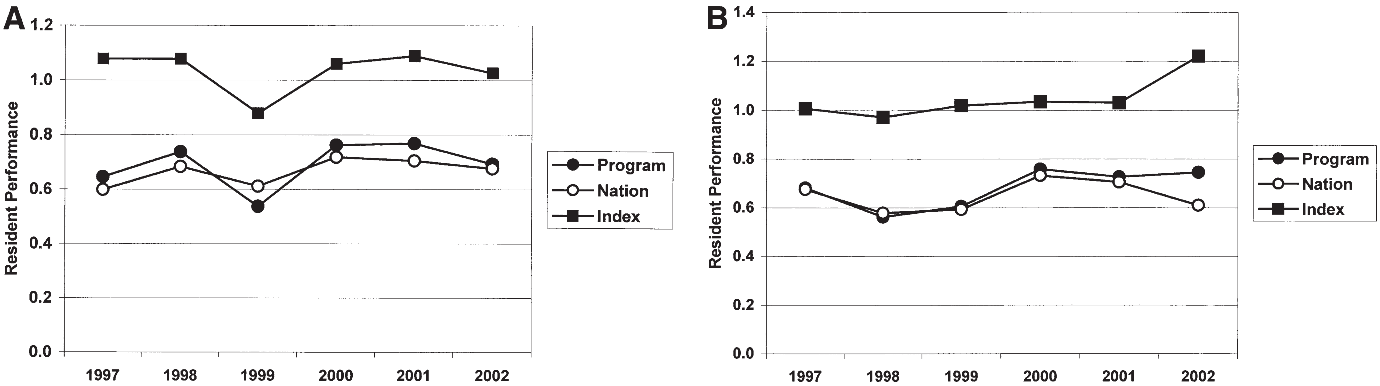

FIG. 2. (A) Sports Medicine, average scores and index. Program average (black circles), national average (white circles), and Sports Medicine score index (squares). Score index corroborates the greatest difference between program and national scores for 1999. (B) Orthopaedic Science, average scores and index. Program average (black circles), national average (white circles), and Orthopaedic Science score index (squares). Score index reveals similar program and national averages until a sharp divergence for 2002.

educators, and many teaching hours were expected after the curriculum had been redesigned at the start of the study period.

Comparing scores by emphasis (score index divided by teaching hours), Orthopaedic Science, Shoulder and Elbow, Rehabilitation, Foot and Ankle, Hand and Spine had high emphasis (Fig. 4B), and three of these six subspecialties also had high score index results for 2002 (Fig. 3). Pediatric Orthopaedics had both a low emphasis and score index. Three of our 4 best score index results had high emphasis whereas 2 of our 4 lowest score index results had low emphasis. Orthopaedic Diseases and Pediatric Orthopaedics were external rotations with expected low emphasis. Orthopaedic Diseases had the second highest score index with the lowest emphasis so its teaching had high efficiency (score index divided by teaching hours, Fig. 4C).

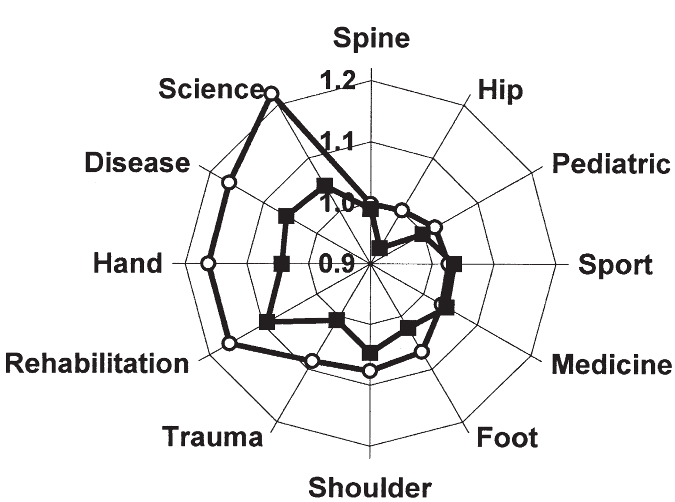

FIG. 3. Subspecialty objective rankings by score index. 2002 (white circles) and 6-year (black squares). The score index benchmark of 1 equals the national average. 2002 score index results are generally better than the 6-year average.

Rankings seemed associated with teaching hours for the subspecialties at the highest and lowest teaching hours, but the association between teaching hours and the number of subspecialty questions was not significant ($P = 0.78$, Fig. 5A). Teaching hours did not reflect the number of subspecialty examination questions. Furthermore, the relationships between resident scores and teaching hours, ($P = 0.71$, Fig. 5B) and between score index and teaching hours were not significant ($P = 0.30$, Fig. 5C).

DISCUSSION

The present educational study demonstrated several methods for assessing graduate medical education. Measuring residency education with specific tools encompassed both teaching and learning, and offered a way to tally academic work. Measuring teaching hours and teaching efficiency quantified academic contribution. We were able to assess our curriculum by comparing subspecialty emphases to a national standard. The analysis detected mismatches of what was taught (i.e., emphasized) and what was tested. Curriculum content adjustments may now be made based on counted teaching hours and scores indexed to the nation. Ranking subspecialties by index results permitted clearer identification of weak and strong subspecialties. Teacher interests and resident needs (low scores) became obvious upon objective ranking. Before this study, surgeon interests were not guided by resident needs in part because we lacked the proper tools to compare student needs (low scores) and teacher interests. Measuring resident score index results permitted statistical testing of educational hypotheses. Because of duty hour restrictions for residents, educators cannot assess curricular content alone but need to assess balances of curricular content, i.e., what is called emphasis in the present study. In a full curriculum with a time restriction, changes must be made cau-

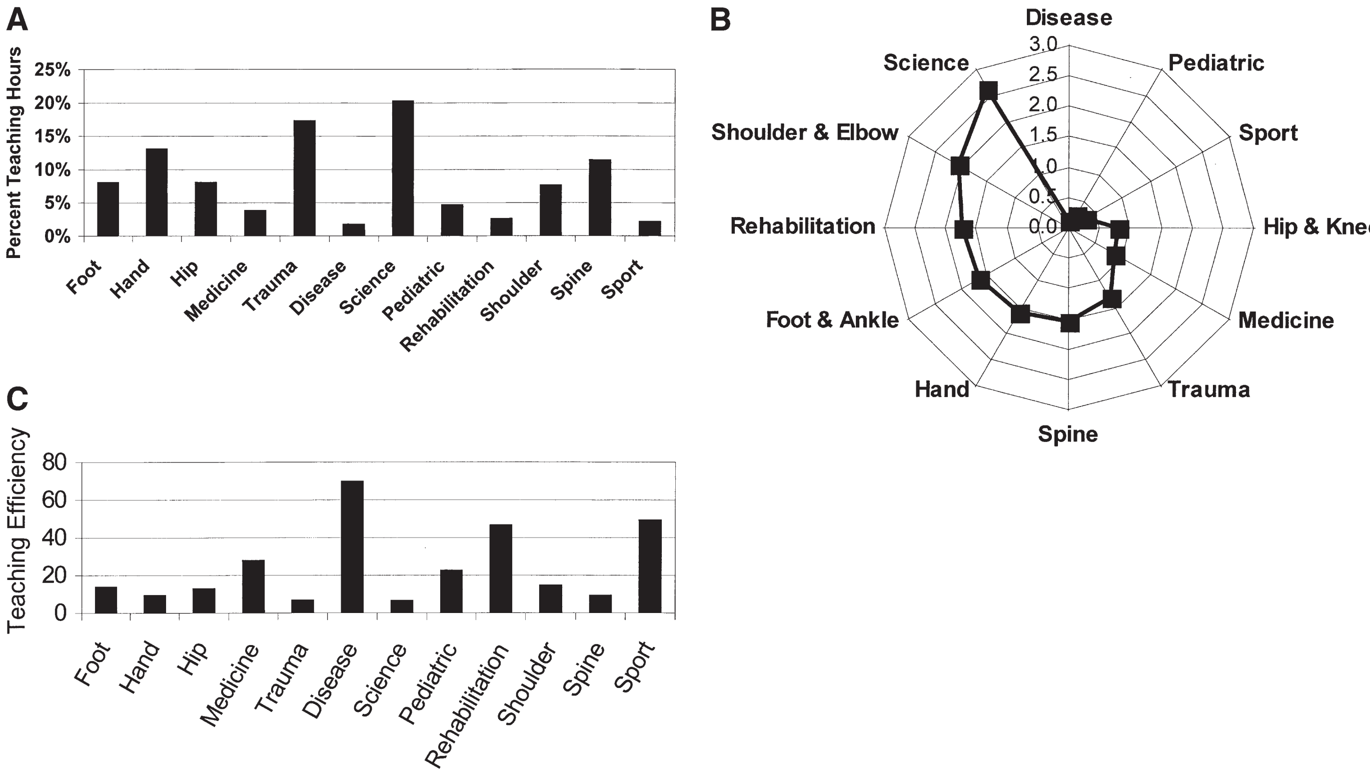

FIG. 4. (A) Percentage of total teaching hours, 2002. (B) teaching emphasis by subspecialty, 2002. An emphasis of 1.0 ± 0.5 meant that the subspecialty teaching hours were balanced with the examination. Nine of 12 subspecialty emphases were unbalanced with the examination. (C) Teaching efficiency by subspecialty in 2002. High efficiency (score index divided by teaching hours) was in subspecialties Orthopedic Diseases, Sports Medicine, and Rehabilitation, whereas low efficiency was in Hand, Spine, Trauma, and Orthopedic Science in descending order. Efficiency ranged from 6 for Orthopedic Science to 69 for Orthopedic Diseases.

tiously as additions means other content must go, making choices a zero-sum gain. As we found no statistical relationship between the quantity of teaching and resident scores, a central issue may a better balance of subspecialty teaching quantities and not be more teaching *per se*. The quality of teaching may be more important than the quantity.

The findings and methods of the present study are novel contributions to medical educational research. The study introduces ways to test statistically educational premises in residency, and reports a quantitative analysis of teaching hours, emphasis, and efficiency. With new, specific, and objective assessment tools for educational effectiveness, the present study used available data in original ways to identify areas for redress or further study. The residency needed a more flexible tool than program average to compare to national data, to rank subspecialty strengths and weaknesses, and to compare yearly trends. The score index had more statistical associations than program average, and was more useful as an assessment tool. The score index assessed residency education thoroughly and permitted a detailed analysis. Trend analysis displayed scoring changes that were undetected prior to score index use. Subspecialty ranking capacity was poor until the score index was used as the score index corrected misconceptions on subspecialty strengths and weaknesses.

An essential ingredient to teacher effectiveness is not subspecialty training but educator interest (both self-declared and manifested by behavior) in teaching the evidence of which we found in various forms. Volunteering extra lectures, surgeon participation at conferences and in resident assessment, and passing on key summaries of current knowledge from subspecialty meetings illustrated interest in teaching relevant to the curriculum. Surgeons taught subjects of their own interest; for example, two surgeons were hand surgeons and gave many hand teaching hours, as hand surgery was of interest to hand surgeons, resulting in a high score index. However, surgeon disinterest in teaching was also evident; surgeons disregarded resident needs evidenced by low scores and taught little outside areas of surgeon interest. Disinterest also was evidenced by not volunteering to fill schedule voids, poor attendance at educational events, late arrival, and early departure from events, rescheduling events because of inadequate preparation, and subject switching regardless of resident need. Resident needs demonstrated by low subspecialty scores did not trump

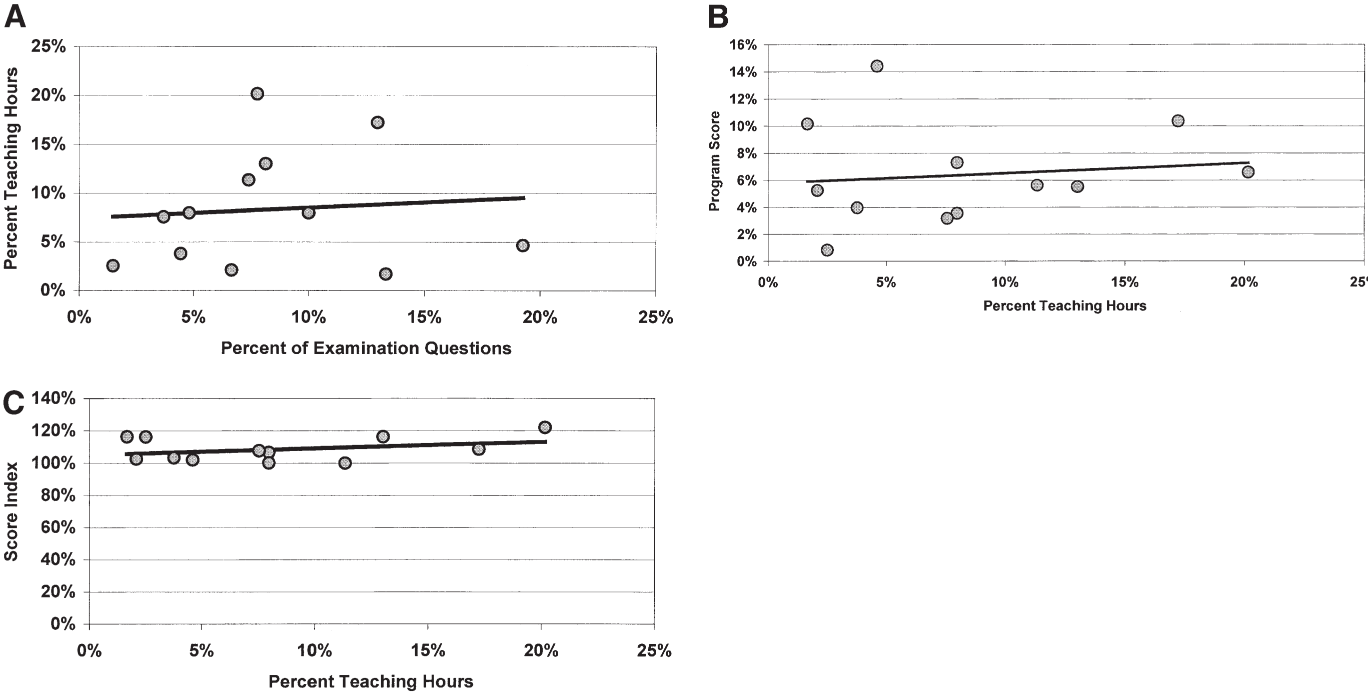

FIG. 5. (A) Relationship of percent teaching hours to percent examination questions, 2002. The regression line is displayed. No significant association was found between the quantities of teaching and testing. (B) Percent program score as a function of percent teaching hours, 2002. The regression line is displayed. No significant association was found between the quantity of teaching and resident scores. (C) Score index as a function of percent teaching hours, 2002. The regression line is displayed. No significant association was found between the quantities of teaching and resident scores.

educator interest so we addressed assertively tension between surgeon interests and resident needs by reinforcing schedules, rewarding compliance, and hosting consultants to address subspecialties in which coverage was weak. Educator behavior and not educator training *per se* was a central issue affecting resident scores.

The strengths of the present study are several. This evidence-based research applied scholarly processes to education. An original design permitted a comprehensive analysis of education with intriguing results of interest to medical educators. Measuring surgeon capacity to teach and contribute academically indicated both academic confidence and accountability. The present work showed residents the value of being effective educators and researchers. The educational process and research caused the hospital administration, the department, the medical educators, the Residency Review Committee evaluator, the medical students, and the house staff to take note. Quantifying educational teaching hours allowed comparison of the curriculum balance to an external standard and yielded the first steps toward an outcomes-based curriculum. Creative methods solved needs.

The weaknesses of the present study are several. The examination has limits in assessing education because it is mainly a knowledge assessment tool and does not measure competence well [9, 15]. The examination may not be the best arbiter of learning nor does it cover all relevant parameters of clinical practice [9]. The examination is an imperfect tool despite its specificity for our residency, its objectivity, and its availability. Our study is norm-referenced to national data and not criterion-referenced to behavior standards [16]. The present study looks more at the curriculum and less at the nature of instruction; for example, the present work did not account for small-group teaching or self-study. The preponderance of teaching is outside of the scheduled program curriculum, the focus of the present study [17]. This focus limits the scope of the study conclusions to curriculum analysis and constitutes a study limitation.

The present study looked more at residency variables than individual person variables although both are important. Residency issues (e.g., teaching hours and emphases) affected scores as did the attendance of some educators. The residency executed a major redesign of the Orthopaedic Science curriculum in 2002 with increased teaching hours. Trend analysis helped discern that the high number of Orthopaedic Science teaching hours seemed more important than individual educator attendance (Fig. 2B). Further, surgeon Z, who was not a sports medicine surgeon, and a Sports Medicine surgeon were involved in a leadership struggle in 1999, and the low scores in Sports Medicine were obvious (Fig. 2A). Surgeon Z was the only surgeon

present only in 1999, and no surgeon was only absent in 1999. Therefore, the high correlation and negative association indicate an important correlation. Obviously, no educator wants to have a negative effect, and the residency variables such as interpersonal issues may be more important than individual person variables such as whether an educator is present or absent. Education, a complex topic, requires study of many important variables beyond the scope of the present work. The general application and validity of novel tools and methods will need further testing by other orthopaedic residencies and specialties, and studies are under way. Although intriguing, application of the findings of the present study should be used with caution until such findings are validated.

In conclusion, the present study measured education by accounting for teaching effectiveness, analyzing resident scoring trends, ranking subspecialty scores, and accounting for teaching hours and emphases. The findings of the present study contribute novel analyses and findings to the growing body of knowledge regarding evidence-based education. The educational tools in the present study helped identify objectively specific areas for redress.

ACKNOWLEDGMENTS

No grant or financial support was used. All authors contributed to design and writing while Drs. Kragh and Ward did the analysis. James A. Farmer, EdD, Professor Emeritus of Continuing Education at the University of Illinois-Urbana, consulted on conception. Lottie B. Applewhite, MS, and Amy W. Newland consulted on writing.

REFERENCES

1. Blumenthal, D., Gokhale, M., Campbell, E. G., and Weissman, J. S. Preparedness for clinical practice: reports of graduating residents at academic health centers. *JAMA.* **286:** 1027, 2001.
2. Bowen, J. L., and Irby, D. M. Assessing quality and costs of education in the ambulatory setting: a review of the literature. *Acad. Med.* **77:** 621, 2002.
3. Luck, J. V., Nestler, S. P., and Simon, M. A. The residency review committee for orthopaedic surgery. *J. Bone Joint Surg. Am.* **83:** 466, 2001.
4. Miller, G. E. Medical education research and development. *JAMA.* **197:** 992, 1966.
5. Bordage, G., Burack, J. H., Irby, D. M., and Stritter, F. T. Education in ambulatory settings: developing valid measures of educational outcomes, and other research priorities. *Acad. Med.* **73:** 743, 1998.
6. Simon, M. A. The education of future orthopaedists-déjà vu. *J. Bone Joint Surg. Am.* **83:** 1416, 2001.
7. Accreditation Council of Graduate Medical Education, ACGME Outcome Project. Available at: http://www.acgme.org/outcome/assess/toolbox.pdf; Internet; accessed August 19, 2004.
8. Mankin, H. J. The orthopaedic in-training examination (OITE). *Clin. Orthop.* **75:** 108, 1971.
9. Nattress L. W. Jr. Orthopaedic in-training examination. *J. Med. Education.* **44:** 880, 1969.
10. Mankin, H. J., Leffert, R. D., Schacar, N. S., Krawcyzk, M., and Barnett, G. O. The use of the orthopaedic educational computer data bank to determine cognitive "profiles" of the orthopaedic in-training examination. *J. Bone Joint Surg. Am.* **60:** 853, 1978.
11. Buckwalter, J. A., Schumacher, R., Albright, J. P., and Cooper, R. R. Use of educational taxonomy for evaluation of cognitive performance. *J. Med. Education.* **56:** 115, 1981.
12. Mankin, H. J., Carter, R. M., and Krawczyk, M. The effect of permissive environment on scoring of the orthopaedic in-training examination. *J. Bone Joint Surg. Am.* **55:** 1100, 1973.
13. McGuire, C. H. The Orthopaedic Training Study-Phase II: the training programs. *Clin. Orthop.* **75:** 49, 1971.
14. Miller, G. E. The orthopaedic training study. *JAMA.* **206:** 601, 1968.
15. Epstein, R. M., and Hundert, E. M. Defining and assessing professional competence. *JAMA.* **287:** 226, 2002.
16. Griffith, C. H., Georgesen, J. C., and Wilson, J. F. Six-year documentation of the association between excellent clinical teaching and improved student examination performances. *Acad. Med.* **75:** S62, 2000.
17. Bassett, C. A. L., and Robinson, R. A. Basic science and research in orthopaedic education. *Clin. Orthop.* **75:** 146, 1971.